Diffusion-based Semantic-Discrepant Outlier Generation for Out-of-Distribution Detection

Suhee Yoon* **Sanghyu Yoon*** **Hankook Lee** **Sangjun Han**
Ye Seul Sim **Kyungeun Lee** **Hyeseung Cho** **Woohyung Lim**
LG AI Research
Seoul, Republic of Korea
{suhee.yoon, sanghyu.yoon, hankook.lee, sj.han,
ysl.sim, kyungeun.lee, hs.cho, w.lim}@lgresearch.ai

Abstract

Out-of-distribution (OOD) detection, which determines whether a given sample is part of the in-distribution (ID) or not, has recently shown promising results by training with synthetic OOD datasets. The important properties for effective synthetic OOD datasets are two-fold: (i) the OOD sample should be close to ID, but (ii) represents distinct semantic information. To achieve this, we introduce a novel framework that consists of **Semantic-Discrepant (SD) Outlier** generation and an improved OOD detection approach with SD outlier. For SD outlier generation, we utilize a conditional diffusion model trained with pseudo-labels. Then, we propose a simple yet effective method, *semantic-discrepant guidance*, allowing model to generate realistic outlier that contains semantically shifted information while preserving nuisances (*e.g.*, background). Furthermore, we suggest SD outlier-aware OOD detector training and scoring methods which improve. Our experiments demonstrate the effectiveness of our framework on CIFAR-10 dataset. We achieve AUROC of 98% when CIFAR-100 are given as OOD. The SD outlier dataset on CIFAR-10 is available at https://zenodo.org/record/8394847.

1 Introduction

Out-of-distribution (OOD) detection is a fundamental machine learning task which aims to detect whether a given sample is drawn from the in-distribution (ID) or not. In decades, OOD detection has found various real-world practical applications, including medical diagnosis [1–3], autonomous driving [4–6], and forecasting [7]. Among a number of OOD detection methods [8–12], one promising approach is to learn a detector using auxiliary OOD dataset, as pioneered by Outlier Exposure (OE) [13]. This makes learning relatively easier since such OOD dataset can provide additional information about discrepancy between ID and OOD. Due to this simple yet effective approach, there have been studied OE-based methods in the recent literature [14–20].

The most challenging part of the approach using auxiliary OOD is dataset acquisition from the real-world. This challenge arises since it is hard to identify whether a sample is OOD and effective enough to learn a detector, especially when label information is unknown. To cope with the challenge, recent research has explored utilizing GANs [21–28] or diffusion models [29] to generate synthetic OOD dataset. To this end, they generate blurred images since such an image can also be considered as OOD. However, these approaches often fail to detect slight semantic differences due to their blurriness and being unrealistic.

The crucial properties for synthetic OOD dataset generation are two-fold: (i) a generated sample should be sufficiently near to ID, but (ii) represents discriminative semantic information. In

*Equal contribution

NeurIPS 2023 Workshop on Synthetic Data Generation with Generative AI.

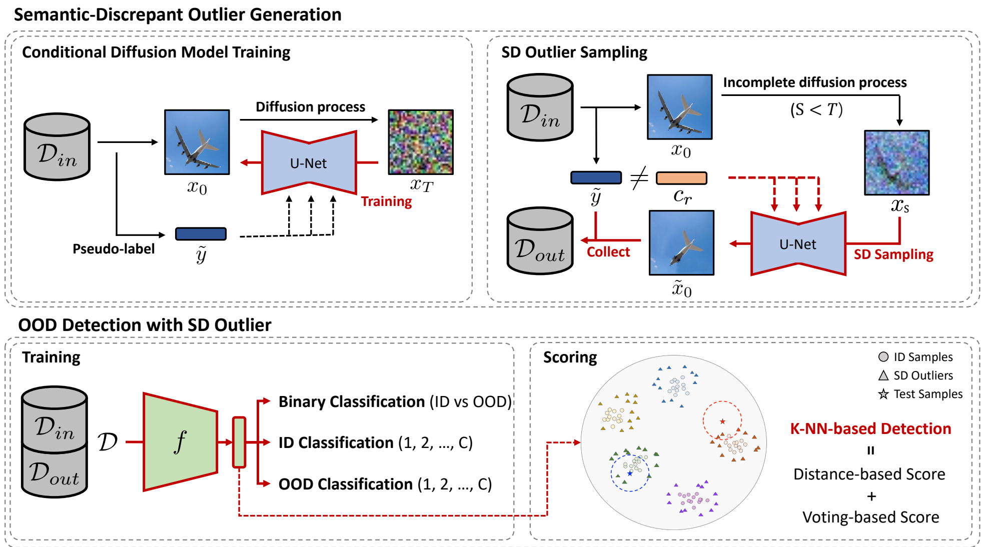


Figure 1: Illustration of our proposed methods framework. This first generates Semantic-Discrepant (SD) Outliers using a conditional diffusion model (top), and then trains a detector using both ID samples and the generated outliers (bottom). For scoring, we use distance-based and voting-based detection scores based on K nearest neighbors on the embedding space.

other words, a sample should be OOD with respect to semantics while preserving nuisances (e.g., background) which have no intrinsic relevance to the semantic. Hence, this OOD dataset offers valuable insights for training detector [30, 16]. Nevertheless, there have been no attempts to address semantically shifted OOD dataset generation.

Contribution. In this paper, we introduce a novel and effective detection framework that consists of (i) **Semantic-Discrepant (SD) Outlier** generation via a diffusion model, and (ii) OOD detection with SD outliers, as illustrated in Figure 1.

- **Semantic-Discrepant Outlier Generation.** Our key idea is generating realistic OOD samples that contains incoherent semantic shift while preserving nuisances with ID. To this end, we propose *semantic-discrepant guidance*, induce the samples struggle with semantic corruption or acquiring odd semantic as shown in Figure 2. We train conditional diffusion model with pseudo-labels obtained from a clustering method (*e.g.*, SCAN [31]). Then we generate SD outlier with semantic-discrepant guidance, which is inconsistent with diffusion process condition. By employing unexpected guidance, the important semantic gradually lose their coherence.
- **OOD Detection with Semantic-Discrepant Outlier.** To develop the utilization of SD outlier, we also suggest SD outlier-aware training and scoring methods for OOD detection. Our loss function encourages the discrimination between ID samples and the SD outliers, while concurrently learns their original semantic information. In addition, we derive OOD scoring function based on both ID and SD outlier samples, distinguishing our approach from the existing method that exclusively relies on only ID.

We demonstrate the effectiveness of the proposed framework through extensive experiments. For example, under OOD generation for unlabeled CIFAR-10, we attain FID score of less than 8, significantly outperforms other generation methods. Moreover, we achieves stats-of-the-art OOD detection performance compared to several baseline methods.

2 Methodology

In this section, we introduce **Semantic-Discrepant (SD) Outlier** generation and its further application of OOD detection framework in detail. In a nutshell, our framework consists of two phases: (a) generating semantic-wise shifted outliers (Section 2.1) and (b) semantic-aware OOD detection with SD outliers (Section 2.2). The overall process is depicted in the Figure 1.

2.1 Semantic-Discrepant Outlier Generation

In this stage, we aim to generate an OOD dataset $\mathcal{D}_{\text{out}}$ that closely resembles $\mathcal{D}_{\text{in}} = \{\mathbf{x}^{(i)}\}$, but contains distinguished semantic information. To achieve this, we first attain semantic controllability by training a conditional diffusion model. Then, we generate SD outliers using *semantic-discrepant guidance* during sampling. Since our generation method is based on diffusion models, we provide its brief background in Appendix A.

Semantic-Aware Diffusion Model Training. Overall, our method is based on a conditional diffusion model under unsupervised learning. To obtain a pseudo-label $\tilde{y}$ for each $\mathbf{x} \in \mathcal{D}_{\text{in}}$, we use a self-supervised method, SCAN clustering [31]. It shows competitive performance with the ground-truth label by successfully generating semantic-wise conditioned samples. This result is consistent with other self-supervised labeling based approaches, *e.g.,* [32]. Hereafter, $\mathcal{D}_{\text{in}} := \{(\mathbf{x}^{(i)}, \tilde{y}^{(i)})\}$ denotes a pseudo-labeled ID dataset.

Classifier-free Diffusion Guidance (CFG) [33] is a simple yet effective conditional diffusion models, avoiding require for a separate classifier. They obtain a combination of a conditional model parameterized with $\epsilon_\theta(\mathbf{x}_t, \mathbf{c})$ and an unconditional model parameterized with $\epsilon_\theta(\mathbf{x}_t) = \epsilon_\theta(\mathbf{x}_t, \mathbf{c} = \varnothing)$, which gives the null token to guidance $\mathbf{c}$ in a single network. During training, it randomly drops the condition with unconditional probability p_{uncond}.

In our method, CFG is trained with $(\mathbf{x}, \tilde{y}) \in \mathcal{D}_{\text{in}}$ as an input. More specifically, the noise prediction model ϵ_θ parameterized by U-Net [34] learns with guidance on the condition $\mathbf{c}_d = (\tilde{y}, t)$ where $\tilde{y}$ is a pseudo-label of x and $t \in [0, T]$ is a diffusion timestep. Thus, the noise prediction model ϵ_θ can be optimized with the following loss:

$$\mathcal{L}(\theta) = \mathbb{E}_{(\mathbf{x}_0, \tilde{y}) \sim \tilde{\mathcal{D}}_{\text{in}}, t \sim \text{Uniform}([0,T]), \epsilon \sim \mathcal{N}(\mathbf{0}, I)}[\|\epsilon_\theta(\mathbf{x}_t, \mathbf{c}_d) - \epsilon\|_2^2]. \tag{1}$$

Semantic-Discrepant Outlier Sampling. Our key idea is *semantic-discrepant guidance*, sampling with c_r selected among other cluster labels, which is inconsistent with the pseudo-label $\tilde{y}$.

$$c_r \sim \text{Uniform}(\{1, ..., C |\, c_r \neq \tilde{y}\}). \tag{2}$$

To be specific, while original method perform sampling from random noise for T timesteps, we utilize $\mathbf{x} \in \mathcal{D}_{in}$ to conduct incomplete diffusion process to a limited timesteps S, which is significantly shorter than T. By stopping diffusion at the early steps, we obtain $\mathbf{x}_S$ with more corruption effect in the semantically important region. While Gaussian noise exhibits a uniform spectral density, the high-frequency components experience more perturbations effect in comparison to the low-frequency components [35]. This distinct property is directly connected to semantic, displaying a correlation with high-frequency components [36, 37]. Then, we start sampling from $\mathbf{x}_S$ with semantic-discrepant guidance $\mathbf{c}_r = (c_r, t)$. We here inject $\mathbf{c}_r$ into the noise prediction function ϵ_t with the timesteps S and feed into every block of the U-Net. Thus, the noise prediction function ϵ_t is rewritten as:

$$\epsilon_t = (1 + w)\epsilon_\theta(\mathbf{x}_t, \mathbf{c}_r) - w\epsilon_\theta(\mathbf{x}_t) \tag{3}$$

Through this process, only semantically contaminated $\tilde{\mathbf{x}}$ is generated as highly diffused regions are contaminated in the direction of $\mathbf{c}_r$ while nuisances are almost preserved their originality. This new SD outlier is saved with the original pseudo-label $\tilde{y}$ into $\mathcal{D}_{\text{out}}$, i.e., $\mathcal{D}_{\text{out}} = \{(\tilde{\mathbf{x}}^{(i)}, \tilde{y}^{(i)})\}$.

2.2 OOD Detection with Semantic-Discrepant Outlier

In this section, we introduce the improved OOD detection strategy to develop SD outlier utilization. The above semantic insights allows semantic aware training and scoring for OOD detection.

Training with SD Outlier. The common approach to learn a detector using auxiliary OOD is multi-task learning. Additionally, we construct a new loss that forcing the model distinguish ID samples from SD outliers while exposing the original semantic to both ID and outlier samples in a different degree. Given $(\mathbf{x}, \tilde{y}) \in \mathcal{D}_{\text{in}} \cup \mathcal{D}_{\text{out}}$ and the dataset indicator function $I(\mathbf{x}) \in \{0, 1\}$, our training objective with SD outlier is formulated as:

$$\mathcal{L}(\mathbf{x}, \tilde{y}) = \mathcal{L}_{\text{CE}}(g_{\text{bin}}(f(\mathbf{x})), I(\mathbf{x})) + (1 - I(\mathbf{x}))\mathcal{L}_{\text{CE}}(g_{\text{in}}(f(\mathbf{x})), \tilde{y}) + I(\mathbf{x})\lambda\mathcal{L}_{\text{CE}}(g_{\text{out}}(f(\mathbf{x})), \tilde{y}) \tag{4}$$

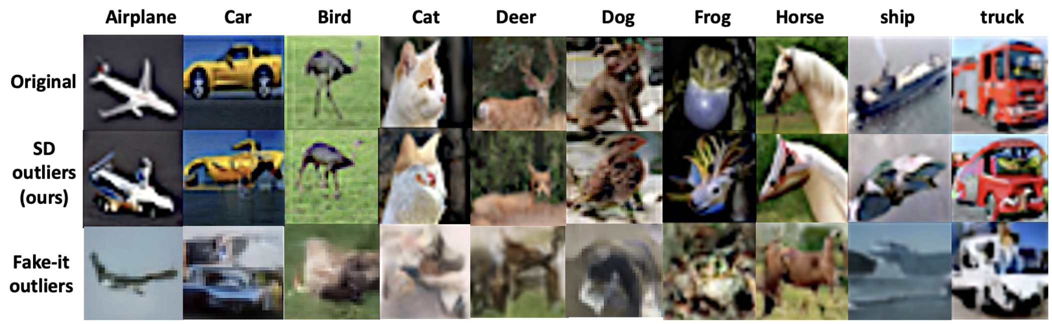


Figure 2: Generated samples on CIFAR-10 (32x32 resolution) comparison with diffusion-based method. As shown, Our SD-outliers are semantically shifted from original samples retaining nuisances while outliers generated from other method [29] are difficult to distinguish.

$$I(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in \mathcal{D}_{\text{in}} \\ 1, & \mathbf{x} \in \mathcal{D}_{\text{out}} \end{cases} \tag{5}$$

The above objective consists of one binary classification g_{bin} and two multi-class classification tasks $g_{\text{in}}, g_{\text{out}}$ while sharing the pre-trained feature extractor f. Binary classification task forces ID samples to separate from outliers. On the other hands, multi-class classification tasks compel samples to discriminate semantics with pseudo-labels. Note that SD outliers have partially semantic contamination, thus we set sufficiently low λ to exposing semantic information in different degree depending on the I. The whole network jointly optimized by minimizing with our loss.

Scoring with SD Outlier. We derive OOD score function based on k nearest neighbours (kNNs). As we attained semantic-wise outliers, we can enhance the scoring methods in two ways, distance-based score and voting-based score. For given input $\mathbf{x}$, its kNNs for distance-based score are denoted as $\{\mathbf{x}_d^1, .., \mathbf{x}_d^k\} \in \mathcal{D}_{\text{in}}$ and for voting-based score are $\{\mathbf{x}_v^1, , .., \mathbf{x}_v^k\} \in \mathcal{D}_{\text{in}} \cup \mathcal{D}_{\text{out}}$. We measure the Euclidean distance between $f(\mathbf{x})$ and $f(\mathbf{x}_d^i)$ and conduct hard voting of $I(\mathbf{x}_v^i)$.

$$\text{OOD Score}(x) = \sum_{i=1}^{k} \left\| f(\mathbf{x}) - f(\mathbf{x}_d^i) \right\|^2 + \alpha \sum_{i=1}^{k} \mathbb{1}(I(\mathbf{x}_v^i) = 1) \tag{6}$$

3 Experiments

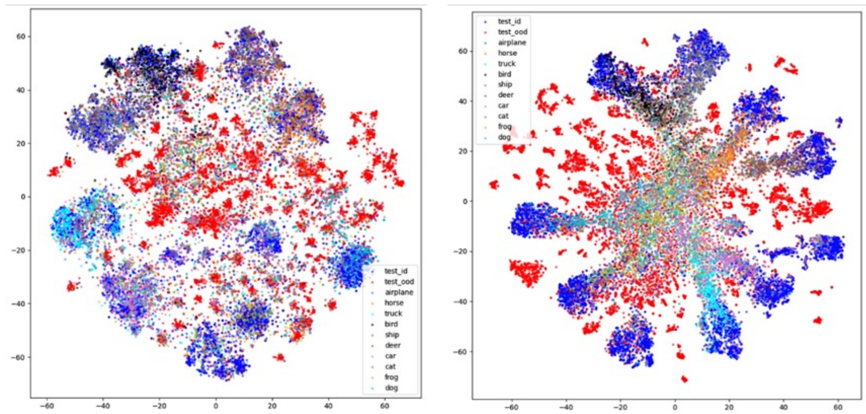
3.1 Setup

We demonstrate our methods on the most commonly used benchmark CIFAR-10. For pseudo-labeling, we use the state-of-the-art clustering algorithm SCAN [31]. We set a number of clusters $C = 10$ which is same as [38] and use default parameter setting from the official implementation. We obtain clustering accuracy 87%.

We train the CFG model using the same architecture as employed on [33]. For our best result, we set diffusion timesteps $S = 100$ and guidance weights $w = 2.0$ for SD outliers sampling. We evaluate the quality of generated outliers with Frechet Inception Distance (FID) with 50000 samples. For our detection network, we adopt ViT-B16 trained with ImageNet as a backbone feature extractor f. We set $\lambda = 0.3$ and $\alpha = 0.3$. More implementation details is in Appendix B. To evaluate OOD detection performance, we use $k = 10$ and Area Under the Receiver Operating Characteristics (AUROC).

3.2 Main Results

Quality Comparison of SD Outlier. Figure 2 shows image samples of original, SD outliers and another diffusion-based method Fake-it [29]. Our samples successfully maintain most of the original image components but cause crucial semantic corruption. For instance, the car loses wheels, the deer sheds antlers, the bird develops four legs, and the dog's head sprouts a beak. Furthermore, ours exhibit a highly realistic appearance with FID score less than 8 while original sampling method shows 2.97 and Fake-it [29] 45.



(a) (b)

Figure 3: Visualization of the embedding space from the feature extractor f by t-SNE. The blue points represent ID (CIFAR-10), red for OOD (CIFAR-100) and SD-outliers are assigned in different colors for each pseudo-label $\tilde{y}$. (a) The embedding space in the early stage of training the OOD detector. (b) The embedding space after training the OOD detector until last epoch.

OOD Detection Performance Comparison. We evaluate the OOD detection performance by testing it on multiple datasets, CIFAR-100, SVHN and LSUN. The most challenging benchmark is CIFAR-100, as it includes the most similar classes found in CIFAR-10.

Table 1 summarizes the OOD detection results compared to previous works on CIFAR-10. Our method outperforms previous methods in all benchmarks dataset. Especially, one notable result is we almost reaching to ground-truth level performance in CIFAR-100 dataset (98.2%). Our approach also surpasses Multi-class AD [38] that solely employs ID samples for both training and scoring.

We visualize the t-SNE plot of the feature space f in Figure 3. We verify that our SD outlier-aware training objective is effective to enhancing OOD detection performance. Figure 3(b) shows that ID samples are discriminated more elaborately with SD outliers. These results demonstrate the effectiveness of the SD outlier dataset in detecting real OOD samples.

Table 1: OOD Detection AUROC (%) on various benchmark datasets. The reported results are over five trials and bold denotes the best results.

	Methods	Networks	(In) CIFAR-10		
			CIFAR-100	SVHN	LSUN
Likelihood	Likelihood	PixelCNN++	52.6	8.3	-
	Likelihood ratio [39]	PixelCNN++	-	91.2	
	Input Complexity [40]	Glow	73.6	95.0	
Self-supervised	Rot [41]	ResNet-18	79.0	97.6	89.2
	GOAD [42]	ResNet-18	77.2	96.3	89.3
	CSI [43]	ResNet-18	89.2	99.8	97.5
	SSD [44]	ResNet-18	89.6	-	-
Pre-trained	DN2 [45]	ResNet-18	83.3	88.9	91
	DN2 [45]	ResNet-152	86.5	96.2	88.7
	MSCL [46]	ResNet-152	90.0	98.6	90.6
	Multi-class AD [38]	ResNet-18	90.8	98.6	98.6
	Multi-class AD [38]	ResNet-152	93.3	99.8	95.4
	Multi-class AD [38]	ViT-B/16	96.7	99.9	99.3
	Fake-it [29]	ViT-B/16	95.7	99.9	99.4
	Ours	ViT-B/16	**98.0**	**99.9**	**99.9**

3.3 Ablation Study

We report ablation studies with CIFAR-10 (ID) vs CIFAR-100 (OOD) comparison.

Dataset Dependence of Sampling Hyperparameter. We confirm that our method is robust to sampling hyperparameters, diffusion timesteps and guidance weight shown on Table 2. Since we are trying to maintain original nuisances, we adopt sufficiently small timesteps $S = \{50, 80, 100, 150, 200\}$. Within this range, we observed that the OOD performance variation is not significant. In addition, we found all state-of-the-art performance on various guidance weight $w = \{2.0, 3.0, 4.0\}$.

Table 2: Ablation study of sampling hyperparameter, timesteps S and guidance weight w.

	Sampling timesteps S				
	50	80	100	150	200
w=2.0	97.6	97.8	**98.0**	97.8	97.4
w=3.0	97.5	97.8	97.9	**98.0**	97.6
w=4.0	97.7	97.7	**98.0**	97.8	97.6

Comparison of OOD Detection Scoring Components We measure the effect of components in our OOD scoring *i.e.,* distance-based score and voting-based score. Table 3 shows the result with $S = 100, w = 2.0$. The most common approach, distance-based score, shows our method has already achieved state-of-the-art compared to other baseline methods. Furthermore, merging with voting score consistently improves in various sampling timesteps S.

Table 3: Ablation study of OOD scoring function, distance-based score and voting-based score.

	Sampling timesteps S				
	50 (FID=4.45)	80 (FID=5.95)	100 (FID=6.88)	150 (FID=7.85)	200 (FID=7.59)
Distance score	97.2	97.5	**97.7**	97.6	97.3
Distance + Voting Score	97.6	97.8	**98.0**	97.8	97.4

Evaluation of SD Outlier Test Dataset. To show the quality of our SD outlier dataset, we evaluate the performance when the generated dataset is considered as test set. The SD outlier test dataset is generated with $S = 150, w = 4.0$ which is different with SD outlier for training dataset $S = 100, w = 1.8$. As we expose SD outlier both training and testing, ours shows best score naturally. However, the other state-of-the-art baselines show severe performance degradation on our dataset. In particular, the multi-class AD method that not exposing any auxiliary outlier degrade to 74.3(%) AUROC score. Therefore, it has been demonstrated that our dataset can become a more challenging semantic-wise OOD detection task.

Table 4: Performance of SD outlier as test dataset on comparable baselines.

Setting	Multi-class AD	Fake-it	Ours
CIFAR-10 vs CIFAR-100	96.7	95.7	**98.0**
CIFAR-10 vs SD outliers	74.3	81.4	**98.2**

4 Conclusion

In this paper, we introduce Semantic-Discrepant (SD) outlier generation and further application to OOD detection framework. Our primary concept is generating realistic OOD samples that semantically shifted while retaining nuisances found in ID by proposing *semantic-discrepant guidance*. Experimental results demonstrate the effectiveness of our approach on several OOD detection benchmarks. It has been proven that our SD outliers can be served as effective auxiliary OOD to learn detector without any additional dataset acquisition efforts. We further can apply our method to large-scale dataset, ImageNet, as [32] surpasses ground-truth performance using pseudo-labels.

References

[1] Zadorozhny K., Thoral, P. Elbers P., and Cinà G. Out-of-distribution detection for medical applications: Guidelines for practical evaluation. *In Multimodal AI in healthcare: A paradigm shift in health intelligence*, pages 137–153, 2022.

[2] Nina Shvetsova, Bart Bakker, Irina Fedulova, Heinrich Schulz, and Dmitry V Dylov. Anomaly detection in medical imaging with deep perceptual autoencoders. *IEEE Access*, 9:118571–118583, 2021.

[3] Stefan Röhrl, Alice Hein, Lucie Huang, Dominik Heim, Christian Klenk, Manuel Lengl, Martin Knopp, Nawal Hafez, Oliver Hayden, and Klaus Diepold. Outlier detection using self-organizing maps for automated blood cell analysis. *arXiv preprint arXiv:2208.08834*, 2022.

[4] Bogdoll Daniel, Maximilian Nitsche, and J. Marius Zöllner. Anomaly detection in autonomous driving: A survey. In *IEEE/CVF conference on computer vision and pattern recognition*, 2022.

[5] Yingda Xia, Yi Zhang, Fengze Liu, Wei Shen, and Alan L Yuille. Synthesize then compare: Detecting failures and anomalies for semantic segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 145–161. Springer, 2020.

[6] Silvio Galesso, Max Argus, and Thomas Brox. Far away in the deep space: Nearest-neighbor-based dense out-of-distribution detection. *arXiv preprint arXiv:2211.06660*, 2022.

[7] Jiawen Xu, Matthias Kovatsch, Denny Mattern, Filippo Mazza, Marko Harasic, Adrian Paschke, and Sergio Lucia. A review on ai for smart manufacturing: Deep learning challenges and solutions. *Applied Sciences*, 12, 2022.

[8] Kimin Lee, Honglak Lee, Kibok Lee, and Jinwoo Shin. Training confidence-calibrated classifiers for detecting out-of-distribution samples. In *International Conference on Learning Representations*, 2018.

[9] Si Liu, Risheek Garrepalli, Thomas Dietterich, Alan Fern, and Dan Hendrycks. Open category detection with pac guarantees. In *International Conference on Machine Learning*, 2018.

[10] Petra Bevandic, Sinisa Segvic, Ivan Kreso, and Marin Orsic. Discriminative out-of-distribution detection for semantic segmentation. In *arXiv:1808.07703.*, 2018.

[11] Andrey Malinin and Mark Gales. Predictive uncertainty estimation via prior networks. In *Neural Information Processing Systems*, 2018.

[12] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations*, 2017.

[13] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. In *International Conference on Machine Learning*, 2019.

[14] Jingyang Zhang, Nathan Inkawhich, Randolph Linderman, Yiran Chen, and Hai Li. Mixture outlier exposure: Towards out-of-distribution detection in fine-grained environments. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, page 5531–5540, 2023.

[15] Cai Jinyu and Jicong Fan. Perturbation learning based anomaly detection. In *Advances in Neural Information Processing Systems*, page 35, 2022.

[16] Jingkang Yang, Haoqi Wang, Litong Feng, Xiaopeng Yan, Huabin Zheng, Wayne Zhang, and Ziwei Liu. Semantically coherent out-of-distribution detection. In *IEEE/CVF international conference on computer vision*, pages 8301–8309, 2021.

[17] Deepak Ravikumar, Sangamesh Kodge, Isha Garg, and Kaushik Roy. Exploring vicinal risk minimization for lightweight out-of-distribution detection. In *arXiv preprint arXiv:2012.08398*, 2020.

[18] Goyal S., Raghunathan A., Jain M., Simhadri H. V., and Jain P. Drocc: Deep robust one-class classification. In *International conference on machine learning*, pages 3711–3721, 2020.

[19] Qing Yu and Kiyoharu Aizawa. Unsupervised out-of-distribution detection by maximum classifier discrepancy. In *IEEE/CVF international conference on computer vision*, pages 9518–9526, 2019.

[20] Sunil Thulasidasan, Gopinath Chennupati, Jeff A Bilmes, Tanmoy Bhattacharya, and Sarah Michalak. On mixup training: Improved calibration and predictive uncertainty for deep neural networks. In *Advances in Neural Information Processing Systems*, page 32, 2019.

[21] Konstantin Kirchheim and Frank Ortmeier. On outlier exposure with generative models. In *In NeurIPS ML Safety Workshop*, 2022.

[22] Hironori Murase and Kenji Fukumizu. Algan: Anomaly detection by generating pseudo anomalous data via latent variables. *IEEE Access*, 10:44259–44270, 2022.

[23] Masoud Pourreza, Bahram Mohammadi, Mostafa Khaki, Samir Bouindour, Hichem Snoussi, and Mohammad Sabokrou. G2d: generate to detect anomaly. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2003–2012, 2021.

[24] Shu Kong and Deva Ramanan. Opengan: Open-set recognition via open data generation. In *IEEE/CVF International Conference on Computer Vision*, pages 813–822, 2021.

[25] Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunjey Choi, and Jaejun Yoo. Reliable fidelity and diversity metrics for generative models. In *International Conference on Machine Learning*, page 7176–7185, 2020.

[26] Daniel Pérez-Cabo, David Jiménez-Cabello, Artur Costa-Pazo, and Roberto J López-Sastre. Deep anomaly detection for generalized face anti-spoofing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[27] Cai Jinyu Lawrence Nea and, Matthew Olson, Xiaoli Fern, Weng-Keen Wong, and Fuxin Li. Open set learning with counterfactual images. In *European Conference on Computer Vision*, page 2018, 613–628.

[28] Jia Deng Jonathan Krause, Michael Stark and Li Fei-Fei. 3d object representations for finegrained categorization. In *IEEE international conference on computer vision workshops*, pages 554–561, 2013.

[29] Hossein Mirzaei, Mohammadreza Salehi, Sajjad Shahabi, Efstratios Gavves, Cees GM Snoek, Mohammad Sabokrou, and Mohammad Hossein Rohban. Fake it until you make it: Towards accurate near-distribution novelty detection. In *The Eleventh International Conference on Learning Representations*, 2022.

[30] Zhang Lily H and Rajesh Ranganath. Robustness to spurious correlations improves semantic out-of-distribution detection. *arXiv preprint arXiv:2302.04132*, 2023.

[31] Wouter Van Gansbeke, Simon Vandenhende, Stamatios Georgoulis, Marc Proesmans, and Luc Van Gool. Scan: Learning to classify images without labels. In *European conference on computer vision*, pages 268–285. Springer, 2020.

[32] Vincent Tao Hu, David W Zhang, Yuki M. Asano, Gertjan J. Burghouts, and Cees G. M. Snoek. Self-guided diffusion models, 2023.

[33] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

[34] Philipp Fischer Ronneberger Olaf and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*, 2015.

[35] Wyatt J. Leach, Schmon S. M., and Willcocks C. G. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 650–656, 2022.

[36] Lee Y., Kim J. Y., Go H., Jeong M., Oh S., and S Choi. Multi-architecture multi-expert diffusion models. In *arXiv preprint arXiv:2306.04990.*, 2023.

[37] et al Wang Haohan. High-frequency component helps explain the generalization of convolutional neural networks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8684–8694, 2020.

[38] Niv Cohen, Ron Abutbul, and Yedid Hoshen. Out-of-distribution detection without class labels. In *European Conference on Computer Vision*, pages 101–117. Springer, 2022.

[39] Jie Ren, Peter J Liu, Emily Fertig, Jasper Snoek, Ryan Poplin, Mark Depristo, Joshua Dillon, and Balaji Lakshminarayanan. Likelihood ratios for out-of-distribution detection. *Advances in neural information processing systems*, 32, 2019.

[40] Joan Serrà, David Álvarez, Vicenç Gómez, Olga Slizovskaia, José F Núñez, and Jordi Luque. Input complexity and out-of-distribution detection with likelihood-based generative models. *arXiv preprint arXiv:1909.11480*, 2019.

[41] Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and Dawn Song. Using self-supervised learning can improve model robustness and uncertainty. *Advances in neural information processing systems*, 32, 2019.

[42] Liron Bergman and Yedid Hoshen. Classification-based anomaly detection for general data. *arXiv preprint arXiv:2005.02359*, 2020.

[43] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. *Advances in neural information processing systems*, 33:11839–11852, 2020.

[44] Vikash Sehwag, Mung Chiang, and Prateek Mittal. Ssd: A unified framework for self-supervised outlier detection. *arXiv preprint arXiv:2103.12051*, 2021.

[45] Liron Bergman, Niv Cohen, and Yedid Hoshen. Deep nearest neighbor anomaly detection. *arXiv preprint arXiv:2002.10445*, 2020.

[46] Tal Reiss and Yedid Hoshen. Mean-shifted contrastive loss for anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 2155–2162, 2023.

[47] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020.